

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 10-285275

(43)Date of publication of application : 23.10.1998

(51)Int.Cl.

H04M 3/00  
G10L 3/00  
G10L 3/00  
H04M 3/42  
H04M 11/00  
H04Q 11/04

(21)Application number : 09-093448

(71)Applicant : NEC CORP

(22)Date of filing : 11.04.1997

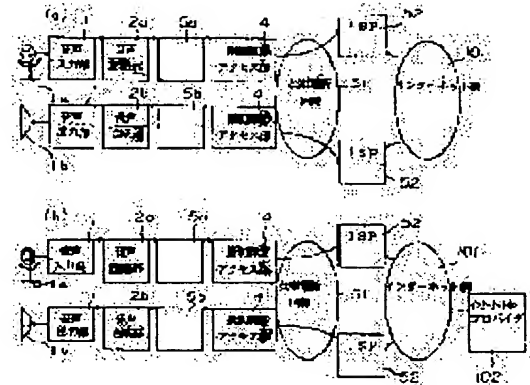
(72)Inventor : KOYAMA AKIRA

## (54) CALLING METHOD, VOICE TRANSMITTING DEVICE AND VOICE RECEIVING DEVICE

## (57)Abstract:

**PROBLEM TO BE SOLVED:** To prevent voice deterioration and the unnatural parting of words and to enable easily comprehensible conversation by converting the voice of a caller into character data, making them into a packet so as to transmit them after that, and outputting received packet-converted character data as voice.

**SOLUTION:** The voice of the caller is inputted to a voice input part 1 through a microphone 1a, converted into voice data, inputted to a voice recognition converting part 2 and converted into a character code data signal by voice recognition. After connection to a reception side terminal, a packet transmitting and communication control part 5a divides the character code data signal by proper length, makes them into the packet and transmits them with a network control and access part 4. Received packet data are demodulated into character code data by the packet and communication control part 5b and, after that, a voice synthesizing and converting part 2 converts them into a voice data signal by voice synthesization. Then, a voice output part 1 outputs the voice data signal as voice through an electric/voice converting means such as a speaker 1b, etc.



## LEGAL STATUS

[Date of request for examination] 11.04.1997

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number] 3237566

[Date of registration] 05.10.2001

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C): 1998,2003 Japan Patent Office

(19) 日本国特許庁 (J P)

(12) 公 開 特 許 公 報 (A)

(11) 特許出願公開番号

特開平10-285275

(43) 公開日 平成10年(1998)10月23日

(51) Int.Cl.<sup>5</sup>

識別記号

F I

H 0 4 M 3/00

H 0 4 M 3/00

B

G 1 0 L 3/00

G 1 0 L 3/00

Q

5 5 1

5 5 1 A

H 0 4 M 3/42

H 0 4 M 3/42

P

11/00

3 0 2

11/00

3 0 2

審査請求 有 請求項の数14 O L (全 12 頁) 最終頁に続く

(21) 出願番号

特願平9-93448

(22) 出願日

平成9年(1997)4月11日

(71) 出願人 000004237

日本電気株式会社

東京都港区芝五丁目7番1号

(72) 発明者 小山 明

東京都港区芝五丁目7番1号 日本電気株式会社内

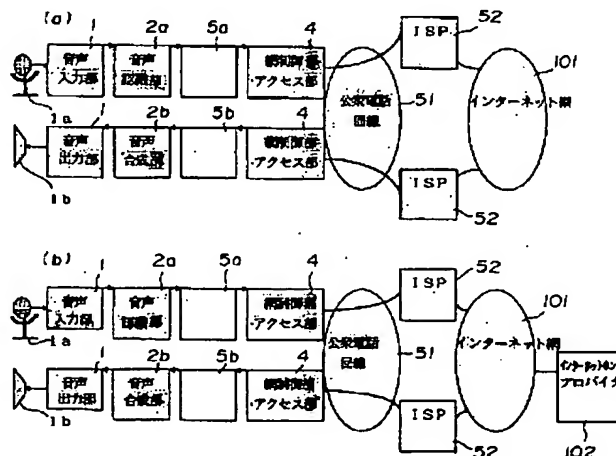
(74) 代理人 弁理士 京本 直樹 (外2名)

(54) 【発明の名称】 通話方法、音声送信装置及び音声受信装置

(57) 【要約】

【課題】 本発明は、非ギャランティー型ネットワークを介した通話方法において、通信の不安定性や負荷変動に起因する音声劣化、言葉の不自然な分断、音飛び等を極力排して理解容易な会話を可能にすると共に、自然な会話を行うことができる、通話方法、音声送信装置及び音声受信装置を提供する。

【解決手段】 話者の音声は、マイクロフォン1a等の音声-電気変換手段を介して音声入力部1に入力され、音声信号がこれに対応する音声データ電気信号に変換される。音声データは音声認識変換部2に入力され、音声認識により文字コードデータ信号に変換される。



5 a : パケット送信および通信制御部  
5 b : パケット受信および通信制御部

## 【特許請求の範囲】

【請求項1】 非ギャランティー型ネットワークを介して話者の音声を通話する方法において、送話側端末において話者の音声を音声データに変換した後、この音声データを音声認識して文字データに変換し、受話側端末の識別データに基づいて前記送話側端末を前記非ギャランティー型ネットワークを介して前記受話側端末に接続し、前記送話側端末と前記受話側端末との接続後、前記文字データをパケット化して前記受話側端末に向けて送信し、前記受話側端末において受信したパケット化文字データを音声合成により音声データに変換した後、音声として出力することを特徴とする通話方法。

【請求項2】 請求項1に記載の通話方法において、前記音声データを所定語数毎に分割して音声認識し、前記文字データに変換することを特徴とする通話方法。

【請求項3】 請求項2に記載の通話方法において、前記音声データを文節毎に分割して音声認識し、前記文字データに変換することを特徴とする通話方法。

【請求項4】 請求項2又は3に記載の通話方法において、前記音声認識に並行して、前記音声データを前記所定語単位で分割してデータ圧縮した圧縮音声データを作成し、前記音声認識が不十分又は不成立の場合は前記圧縮音声データをパケット化して送出することを特徴とする通話方法。

【請求項5】 請求項1～4のいずれかに記載の通話方法において、前記送話側端末に入力される音声の特徴を示す音声プロフィールをこの送話側端末から前記受話側端末に予め送信しておき、前記受話側端末において、音声合成時に前記音声プロフィールに基づいて前記文字データを前記音声データに変換することを特徴とする通話方法。

【請求項6】 請求項1～4のいずれかに記載の通話方法において、前記送話側端末及び前記受話側端末の双方において、標準音声に対する音声合成標準パターンを用意しておき、前記送話側端末において、この送話側端末に入力された音声と前記標準音声との差分データを作成して前記パケット化文字データとともに送出し、前記受話側端末において、前記音声合成標準パターン及び前記差分データに基づいて音声合成することを特徴とする通話方法。

【請求項7】 請求項1～6のいずれかに記載の通話方法において、前記送話側端末において背景音に対応する背景音源データを前記パケット化文字データとともに送出し、前記受話側端末において、前記音声データと前記背景音とを重畳して出力することを特徴とする通話方法。

【請求項8】 請求項1～7のいずれかに記載の通話方

法において、

前記送話側端末に、サンプル音声データを予め用意しておき、

前記送話側端末に前記話者から最初の音声入力があった際に前記音声データと前記サンプル音声データとを比較して、前記音声データと前記サンプル音声データとが同一話者によるものであるか否かを確認し、同一話者によるものであることが確認されたときにのみ通話を許可することを特徴とする通話方法。

10 【請求項9】 請求項1～8のいずれかに記載の通話方法において、

送話側端末において、前記話者を撮像した画像データを作成してこれを受話側端末に送出し、

受話側端末において、受信した前記画像データに基づいて画像を表示することを特徴とする通話方法。

【請求項10】 請求項1～9のいずれかに記載の通話方法において、

受話側端末において、受信したパケット化文字データを文字形式で表示することを特徴とする通話方法。

20 【請求項11】 請求項1～10のいずれかに記載の通話方法において、

受話側端末において、受信したパケット化文字データを電子メール形式に変換することを特徴とする通話方法。

【請求項12】 話者の音声を音声データに変換する音声入力手段と、

前記音声データを音声認識して文字データに変換する音声認識手段と、

非ギャランティー型ネットワーク上の特定の通話相手を識別するための識別データを入力する識別データ入力手段と、

30 前記識別データに基づいて前記非ギャランティー型ネットワークを介して前記特定の通話相手に接続するネットワーク接続手段と、

前記文字データをパケット化して非ギャランティー型ネットワークに送出する送信手段とを備えたことを特徴とする音声送信装置。

【請求項13】 請求項12記載の音声送信装置において、前記音声認識手段は、音声を認識して前記文字データに変換する際、前記文字データの変換に成功した場合には変換に成功した前記文字データをパケット化し、認識に失敗した場合には認識に失敗した前記音声データを圧縮符号化してパケット化することを特徴とする音声送信装置。

【請求項14】 自身を特定する識別信号に基づいて非ギャランティー型ネットワークと接続するネットワーク接続手段と、

前記非ギャランティー型ネットワークからパケット化された文字データを受信する受信手段と、

50 受信された前記パケット化文字データを音声合成により音声データに変換する音声合成変換手段と、

## 3

前記変換された音声データに基づいて音声を出力する音声出力手段とを備えたことを特徴とする音声受信装置。

## 【発明の詳細な説明】

【発明の属する技術分野】本発明は、インターネット等の非ギャランティー型ネットワークを介して通話を行う通話方法、音声送信装置及び音声受信装置に関する。

## 【0001】

【従来の技術】インターネットの普及に伴い、このインターネット網を介して音声を相互に伝送して公衆電話回線と同様に双方向の通話を行う技術が提案されている。このような技術はインターネットホンをサービスと呼ばれる。

【0002】現在のインターネットホンは、コンピュータとルータによるネットワークを多段に経由するため、各接続点でのサーバの負荷による遅延の影響を受け、遅延と言葉の不自然な分断、音飛び等が発生し完全な通話が出来ないのが普通である。これは、インターネットは情報の時間、到達を保証しない非ギャランティー型ネットワーク（ベストエフォート型ともいう）であることに要因がある。

【0003】この点を解決するために、リアルタイムプロトコルや回線を確保するリザーブプロトコルが提案されているが、基本的に多数の通話者がネットワークを共用する事からお互いの完全な意図疎通を保証するに至っていない。

【0004】上述の問題の原因にインターネット自体の伝送能力が深く関わることは当然であるが、音声データの圧縮能力の問題も関わっている。音声データの圧縮には、圧縮率を上げると音声品質が劣化し、圧縮率を低く押さえると音声品質は上がるが伝送路の帯域を大幅に消費する事から遅延が生じ音声の取りこぼし音飛び・空白等が生じて、いずれのケースでも所望の通話品質が得られない。

【0005】狭帯域で多くの音声情報を送る手法として、入力された通話者の音声音声認識技術により文字データに変換し、文字データを受話側に送信する技術がある。音声情報に比較して文字データの情報は格段に少ないので、通信遅延を短くすることができ、加えて音声圧縮に伴う問題も生じない。このような技術の一例としては、特開昭60-136450号公報、特開昭61-256848号公報が従来知られている。

【0006】前者はパケット交換機の端末間で、入力音声認識した後該交換機のパケット形式のデータに変換し伝送する方式を示しており、交換機の処理効率を向上させる手段として提案されている。本方式では、単一のパケット交換網内における伝送のため、多くの未知の接続点を経て伝達されるインターネット網で生ずる大幅な遅延や、負荷変動に対する対処はされておらず、また、同公報にも記載されているように、画一の音色での再生にとどまっており自然な音声への配慮はなされていない

## 4

等フレキシブル且つ広範囲の情報伝送への適用は困難である。

【0007】後者は、送話端末で音声認識を行い文字コードを交換機を通じて相手端末に送り、受話端末で音声合成を行う構成を示している。本方式も、単一の通信回線内における伝送のため、上述した前者の方式と同様に、多くの未知の接続点を経て伝達されるインターネット網で生ずる大幅な遅延や、負荷変動に対する対処はされておらず、また、リアルタイムあるいは自然な音声への変換の工夫は成されていない。

## 【0008】

【発明が解決しようとする課題】従って、上述した従来の技術には次のような課題が存する。

【0009】第1の課題は、インターネット網自体の伝送能力、多段に構成されているサーバ等未知の通信路を経るための通信の不安定性や負荷変動に起因する、音声の劣化と言葉の不自然な分断、音飛び等により意味の理解できる会話が出来にくいことである。

【0010】第2の課題は、音声認識を用いた音声伝送システムにおいても、文字コードによる伝達のみでは、機械的な音声再生されるため自然な会話が出来にくく意味の取り違いなどの問題が発生する可能性があることである。

【0011】本発明は上述した課題に鑑みてなされたもので、その目的とするところは、非ギャランティー型ネットワークを介した通話方法において、通信の不安定性や負荷変動に起因する音声劣化、言葉の不自然な分断、音飛び等を極力排して理解容易な会話を可能にすると共に、自然な会話をする事ができる、通話方法、音声送信装置及び音声受信装置を提供する点にある。

## 【0012】

【課題を解決するための手段】本発明は、非ギャランティー型ネットワークを介して話者の音声を通話する方法に適用される。

【0013】そして、上述の課題は、送話側端末において話者の音声音声データに変換した後、この音声データを音声認識して文字データに変換し、受話側端末の識別データに基づいて送話側端末を非ギャランティー型ネットワークを介して受話側端末に接続し、送話側端末と受話側端末との接続後、文字データをパケット化して受話側端末に向けて送信し、受話側端末において受信したパケット化文字データを音声合成により音声データに変換した後、音声として出力することにより解決される。

【0014】本発明によれば、次のようなことが達成できる。まず第1に、認識によるデータ量の削減により通信の遅延を無くすることができる。話言葉の文節毎に意味を解釈し、文字データというデータ量の非常に少ない形式に変換してしまう為、通信の遅延を非常に短くすることが出来る。この為、早いタイミングで正しい情報を得ることが出来る。

【0015】第2に、認識による意味情報の相手への確実な伝達が可能になる。話し言葉を認識し、まとまった文字情報にする事により、音というノイズの影響を受け易い曖昧な情報でなく明瞭に意味のある言葉として再生する事により話者の意志を正しく伝えることが出来る。

【0016】この際、相手に正しい情報を伝えるためには、多少の遅延があっても正確な文章がやりとりされる必要がある。従来方式では、言葉の内容に依存しないでコード化していた為通信の遅延、とぎれ等があると意味が全く通じなくなる恐れがあった。この為、音声により、わかりやすく確実に伝えるための手段として、下記のような構成を採用してもよい。

【0017】すなわち、上述の通話方法において、音声データを所定語数毎に分割して音声認識し、文字データに変換してもよい。この場合、音声データを文節毎に分割して音声認識し、文字データに変換してもよい。

【0018】また、上述の通話方法において、音声認識に並行して、音声データを所定語単位で分割してデータ圧縮した圧縮音声データを作成し、音声認識が不十分又は不成立の場合は圧縮音声データをパケット化して送出するようにしてもよい。

【0019】さらに、送話側端末に入力される音声の特徴を示す音声プロフィールをこの送話側端末から受話側端末に予め送信しておき、受話側端末において、音声合成時に音声プロフィールに基づいて文字データを音声データに変換するようにしてもよい。あるいは、送話側端末及び受話側端末の双方において、標準音声に対する音声合成標準パターンを用意しておき、送話側端末において、この送話側端末に入力された音声と標準音声との差分データを作成してパケット化文字データとともに送出し、受話側端末において、音声合成標準パターン及び差分データに基づいて音声合成するようにしてもよい。

【0020】さらに、送話側端末において背景音に対応する背景音源データをパケット化文字データとともに送出し、受話側端末において、音声データと背景音とを重畳して出力してもよいし、送話側端末に、サンプル音声データを予め用意しておき、送話側端末に話者から最初の音声入力があった際に音声データとサンプル音声データとを比較して、音声データとサンプル音声データとが同一話者によるものであるか否かを確認し、同一話者によるものであることが確認されたときにのみ通話を許可するようにしてもよい。送話側端末において、話者を撮像した画像データを作成してこれを受話側端末に送出し、受話側端末において、受信した画像データに基づいて画像を表示してもよい。受話側端末において、受信したパケット化文字データを文字形式で表示し、また、受話側端末において、受信したパケット化文字データを電子メール形式に変換してもよい。

【0021】また、本発明は、話者の音声を音声に関するデータに変換して送信する音声送信装置に適用され、

そして、上述の課題は、話者の音声を音声データに変換する音声入力手段と、音声データを音声認識して文字データに変換する音声認識手段と、非ギャランティーネットワーク上の特定の通話相手を識別するための識別データを入力する識別データ入力手段と、識別データに基づいて非ギャランティー型ネットワークを介して特定の通話相手に接続するネットワーク接続手段と、文字データをパケット化して非ギャランティー型ネットワークに送出する送信手段とを設けることにより達成される。

【0022】音声認識手段によって音声を認識して文字データに変換する際、文字データの変換に成功した場合には変換に成功した文字データをパケット化し、認識に失敗した場合には認識に失敗した音声データを圧縮符号化してパケット化することもできる。

【0023】また、本発明は、音声に関するデータを受信して音声に変換する音声受信装置に適用され、そして、上述の課題は、自身を特定する識別信号に基づいて非ギャランティー型ネットワークと接続するネットワーク接続手段と、非ギャランティー型ネットワークからパケット化された文字データを受信する受信手段と、受信されたパケット化文字データを音声合成により音声データに変換する音声合成変換手段と、変換された音声データに基づいて音声を出力する音声出力手段とを設けることにより達成される。

【0024】

【発明の実施の形態】次に本発明の実施形態について、図面を参照して説明する。

【0025】一第1実施形態一

図1(a)は、本発明の通話方法が適用される音声送信装置及び受信装置の第1実施形態を示すブロック図である。

【0026】この図において、話者の音声は、マイクロフォン1a等の音声-電気変換手段を介して音声入力部1に入力され、音声信号がこれに対応する音声データ電気信号に変換される。音声データは音声認識変換部2に入力され、音声認識により文字コードデータ信号に変換される。たとえばモデム又はターミナルアダプタからなる網制御・アクセス部4は、公衆電話回線51、インターネットサービスプロバイダ52を経由しインターネット101に接続する。この網制御・アクセス部4が通話相手の受信側端末の網制御・アクセス部4とインターネットプロトコルにより接続を確立した後、音声認識により通話に入る。この際、送信側端末は受信側端末の識別番号、一般的にはIP(Internet Protocol)アドレスを事前に知っており、このIPアドレスを用いて受信側端末との接続を確立する。

【0027】受信側端末との接続確立後、パケット送信および通信制御部5aは文字コードデータ信号を適当な長さに区切ってパケット化して、IPアドレス、データ番号等を付して網制御・アクセス部4を介して受信側

端末に伝送する。受信側端末では受信したパケットデータをパケット受信および通信制御部5bで文字コードデータ信号に復調した後、音声合成・変換部2で音声合成により音声データ信号に変換する。そして、音声出力部1は、音声データ信号をスピーカ1b等の電気-音声変換手段を介して音声として出力する。

【0028】図1(b)は、本発明の通話方法が適用される音声送信装置及び受信装置の第1実施形態の変形例を示すブロック図である。

【0029】この変形例では、インターネット網101内のインターネットホンスービスプロバイダ102に送信側、受信側端末がそれぞれ接続されている状況において、送信側が出した相手識別番号をインターネットホンスービスプロバイダ102が受け取り、インターネットホンスービスプロバイダ102が受信側への接続プロセスを実行した後、端末間の音声認識による通話を開始する。

【0030】本接続は、相手側端末がダイヤルアップ接続によりインターネットサービスプロバイダ52に接続している場合のように、相手側端末が固定の識別番号(IPアドレス)を持たない場合に適用される。この際、それぞれの端末側で事前にインターネットホンスービスプロバイダ102に自分の電話番号やメールアドレスなど接続に必要な識別ID情報を登録しておき、送信側端末が受信側端末の識別ID情報を入力するとこの識別ID情報からインターネットホンスービスプロバイダ102がモデム、ターミナルアダプタ等のアクセスデバイスを介し受信側端末に発呼することにより受信側端末への接続を確立する。あるいは、双方の端末が同時期にインターネットホンスービスプロバイダ102に接続している場合にも接続確立が可能である。

【0031】図1に示す装置は、パーソナルコンピュータ(PC)等の処理装置に通信装置を付加し組み合わせで構成するタイプと、処理部、通信部を一体としたタイプのいずれか又はそれぞれの特長を合わせたタイプが可能である。

#### 【0032】—第2実施形態—

図2は、本発明の通話方法が適用される音声送信装置及び受信装置の第2実施形態を示すブロック図である。

【0033】本装置、すなわち送信側端末20a及び受信側端末20bはそれぞれ、マイクロフォン1a、スピーカ1bを内蔵する音声入出力部(音声入力手段、音声出力手段)1と、音声入出力部1からの音声データを音声認識し文字コードデータに変換する音声認識・合成変換部(音声認識手段、音声合成変換手段)2と、通信制御手順の全体を制御する通信制御部(送信手段、受信手段)3と、網制御装置(ネットワーク接続手段)4と、モデム5と、各種データが入力される入力操作部6と、音声又はデータ蓄積・表示部7を備える。

【0034】音声認識・合成変換部2は、図3に示すよ

うに、音声認識回路2aと、圧縮符号化回路2bと、選択回路2cを有する。音声認識回路2aは、音声認識された音声データを予め記憶された登録辞書の用語と比較することで「もしもし」のような音節単位に認識し、認識結果を文字コードとして選択回路2cに出力する。また、認識が成功したか失敗した(登録辞書に登録された音声ではなかった)のかを示す信号が選択回路2cに出力される。圧縮符号化回路2bは、音声信号を圧縮符号化し選択回路2cに出力する。選択回路2cは、通常、音声認識回路2aの出力を選択しており、認識が失敗したことを示す信号が供給されると、圧縮符号化回路2bの出力を選択する。

【0035】送信側端末20aの入力操作部6から入力された受信側端末20bの識別番号に基づき、電話回線51及びインターネットサービスプロバイダ(ISP)52を経由し送信側端末20aがインターネット網101に接続される。そして、同様の手順によりインターネット網101に接続された相手端末20bとインターネットプロトコルにより接続を確立した後、送話時には、音声認識・合成変換部2にて、音声を認識し文字コードに変換し、この文字コードをパケット化してIPアドレスを付し受信側端末20bにTCP/IPで送り、相手端末20bにおいては文字コードを音声合成により音声に変換する事により音声を相手に伝え、受話に際しては相手端末20bから送られたパケットデータを送信側端末20aで順次音声に合成し、話者は互いに連続した文章で会話できる。

【0036】本構成を有する通信装置の動作を同図を用いて詳しく説明する。送信側端末20aの音声入出力部1はマイクロフォン1aから入った音声をまずアナログからデジタルに変換し、音声認識・合成変換部2に送る。音声認識・合成変換部2においては、音声パターン及び登録辞書を用いて入力音声を文字コードに変換する。入力が完了したら通信制御部3は直ちに、文字コードをパケット化しIPアドレスを付してモデム5、網制御装置4、通信回線を経て、相手端末に転送する。

【0037】ここで音声は単語毎にされ、識別した内容が極力遅延無きよう相手端末に送られるが、互いに接続されたインターネットサービスプロバイダ52、インターネット網101を経由して相手に届く為こうした網の遅延時間が影響し相手には1呼吸以上遅れて到着する。

【0038】尚、本図中音声認識・合成変換部2は、音声テキスト変換の他にフォーマット変換あるいは圧縮符号化等の機能も有する物とする。

【0039】図4は図2に示す音声送信装置・音声受信装置の具体的な動作を示すフローチャートである。

【0040】まず、発呼側(送信側)端末20aからまずダイヤルアップ等によるインターネット呼接続を着呼側(受信側)端末20bに対して行う。

【0041】インターネット呼接続とは、まず発呼側端

末20aの網制御装置4が公衆電話回線51経由でインターネットサービスプロバイダ52にログオンしインターネット網101に接続する。

【0042】次いで、通信制御部3から網制御装置4、モデム5を経由してインターネットホスサーバ102に接続し、つなぎたい相手、すなわち着呼側端末20bの識別番号を送出する。

【0043】インターネットホスサーバ102は識別番号と相手IPアドレスを変換する表を保有しており着呼側端末20bにIPアドレスを利用して接続要求する。着呼側端末20bは、その制御のもとにあらかじめダイヤルアップ等によりインターネット網101にログオンし同一のインターネットホスサービスプロバイダ102に接続するか、あるいはインターネットホスサービスプロバイダ102が着呼側端末20bに最も近いアクセス点を経由し電話接続して、着呼側端末20bが自動応答等により電話呼接続とインターネット網接続を完了しているものとする。

【0044】インターネット呼接続は、上述のように中間にインターネットホスサービスプロバイダ102のようなIPアドレスと個人名を読み替える機構を介するか、あるいは相手が固定のIPアドレスを持っており発呼側がそのアドレスをあらかじめ知っていて、そのIPを送出するかいずれかの方法により行う。

【0045】以上の手順は図4のステップS101、S201において実行される。

【0046】インターネット呼接続完了後、以下に示す手順により会話を行う。

【0047】相手との会話は、端末20a、20bの音声認識・合成変換部2の音声認識、圧縮伸張機能を使って行う。この場合音声認識が主であり、圧縮伸張機能は予備として機能する。

【0048】会話に先立って発呼側端末20aの通信制御部3が相手側の音声合成再生時に、発呼者の発音に近い音色で再生されるようにするためのデータ、音声プロフィールを送出するか相手端末20bとのプロトコルにより確認する。確認の結果、“送る”が選択された場合には、発呼側端末20aの標準又は特に音色を変えたい場合には特に指定した音声プロフィールを相手端末20bに送出する。同様に受話側端末20bからも音声プロフィールの送信の確認が行われ、“送る”が選択された場合には、受話側端末20bから音声プロフィールが送られ送話側端末20aは受信の動作を行う。送られた音声プロフィールは受話・送話双方で組み込みの操作が行われる。音声プロフィールとは、話者本人の声色などを特徴づける、文字から音声に変換する際の規則（例えば男声、女声など）を音波のデジタルパターンで示したものであり、送信側端末20aにおいて事前に学習されたものであることから、受話側端末20bの音声認識・合成変換部2においてこのパターンを組み込み送られた文

字コードデータを音声合成すると相手の声色に近い再生音を得られる。

【0049】以上の手順は図4のステップS102～S105及びS202～S205において実行される。

【0050】次に、実際の会話に入った段階の送信側端末20aの手順をステップS106～S110に示す。

【0051】会話においては、以下の操作が繰り返され相手に言葉が伝わる。話された音声に対し、音声認識・合成変換部2の音声認識回路2aにおいて文節単位の音声データの切り出し、音声認識及び認識の成否確認を行う。成功の場合には認識後の文字コードデータが選択回路2cを介して着呼側端末20bにパケットデータとして送られる。音声認識は100%の成功率ではないため、発声不明瞭などの理由により認識失敗のケースもある割合で発生する。失敗の場合に備え、音声認識と同一の位置で切り出された圧縮データも圧縮符号化回路2bで並行して作成され、音声認識に失敗した場合は認識されたデータではなく圧縮データで失敗した部分を埋める。

【0052】音声の「切り出し」は、通常の音節単位で行う。即ち“こんにちは”、“私です”等の単位で順次認識する。

【0053】次いで音声認識回路2aは、発生音声登録辞書と比較し、類似度の大きいパターンを採用する。比較して類似語のないデータは新語か認識できない語のため、圧縮符号化回路2bで作成された音声圧縮データを選択回路2cが前記の通り埋め込む。

【0054】次に、実際の会話に入った段階の着呼側端末20bの手順をステップS210～S214に示す。

【0055】着呼側端末20bにおいては、発呼側端末20aから送られてくるパケットデータをチェックし、テキスト形式、すなわち文字コードデータの場合は、音声認識・合成変換部2において既に組み込み済みの音声プロフィールに基づく音声パターンを用い音声データに変換しスピーカ1bを有する音声入出力部1に送出する。圧縮データの場合は音声認識・合成変換部2の伸張機構を使い入力音声相当の音声データに変換し音声入出力部1に送出する。

【0056】次に、図5を用い、話した言葉がどのように受信側機器で再生されるかを説明する。図5はデータ処理の時間的関係を示すタイムチャートである。

【0057】例えば、“もしもし田中です。田代さんをお願いします”という言葉の場合には、4つの音節“もしもし”“たなかです”“たしろさんを”“おねがいます”となり、これらを順次認識しテキストとして送出、インターネット・LAN等の伝送路を経由し受信側に送信される。受信側では、受信信号を順次合成して音声データとし、音声出力機器を経由し話者に音声を伝える。

【0058】(a)に示す正常時は、上記手続きにて音



声の再生が行われるが、(b)に示すように認識しにくい言葉があり認識に失敗した場合は、予め並行して作成されている圧縮データをテキストデータの代わりに受信端末に送出、受信端末において伸張の処理を行い合成によって得られた他の言葉と順序関係を合わせ再生される。

【0059】本実施形態は以上のように構成されているので、以下に掲げる効果を奏する。

【0060】すなわち、本実施形態では、話言葉の文節毎に意味を解釈し、文字データというデータ量の非常に少ない形式に変換してしまう為、通信の遅延を非常に短くすることが出来、且つ文字は単なる音の連続でなく明瞭に意味のある言葉として再生される。この為、通信の不安定性や負荷変動に起因する音声劣化、言葉の不自然な分断、音飛び等を極力排して理解容易な会話を可能にすると共に、自然な会話をすることができるまた、圧縮機能を追加した音声認識との複合機能は、従来考えられている音声認識の入力音声の不明瞭さや処理能力に起因する認識性能・信頼度の問題を解決し、市販装置への適用を実現し易くする事が可能である。

【0061】さらに、文字で蓄積転送できることにより、インターネット等の各種ネットワークの特性に応じた呼蓄積やメールとしての転送、同報、画像との組み合わせへの応用が可能である。

【0062】当然ながら、トータルで伝送するデータ量が少ないと言うことは、インターネットサービスプロバイダの能力への影響も少なく、広い範囲での利用が可能な他、インターネットトラフィックを下げ網全体の活用幅を広げる社会的効果もある。

#### 【0063】一第3実施形態一

本発明の通信方法が適用される音声送信装置・受信装置の第3実施形態を、図2及び図4を用いて説明する。

【0064】図2に示す構成の装置において、図4のステップS101、S201におけるインターネット呼接続の段階で、音声認識・合成変換部2により最初に発声する代表的な言葉“もしもし”等の発声音声波形を分析し、予め登録した本人の音声パターンと比較する事により本人であることを確認し、本人であることが確認された時にのみ以降の通話を可能としたり、あるいは認証結果に基づいて登録音声パターンを自動選択して相手に送出する。この他に、音声認識・合成変換部2による確証結果を元に通話以外の各種サービスへのアクセスを可能とするなどの多彩なサービスが可能である。

#### 【0065】一第4実施形態一

本発明の通信方法が適用される音声送信装置・受信装置の第4実施形態を、図2及び図6を用いて説明する。

【0066】第1実施形態では、音声合成に際し、図6(a)に示すように本人の特徴を反映した辞書(音声プロファイル)を予め送出しておくことにより、基準辞書を本人の音声に合わせてチューニングし本人の声に近い

形で再生している。この実施形態では、図6(b)に示すように基準辞書は組み込まれている標準のものを使い、文字コード作成の際の標準パターンとの差分データを添付して送出することにより、個々の音節のデータを本人の声色に近づけるアルゴリズムを追加することにより、最初の音声パターンの送出を省略することが可能である。

#### 【0067】一第5実施形態一

本発明の通信方法が適用される音声送信装置・受信装置の第4実施形態を、図2を用いて説明する。この実施形態では、上述の音声プロファイルとは別に背景音源コードを文字データに添付して送ることにより、背景音を会話に合成に際して再生することが可能である。すなわち、上述の第1実施形態では音声認識後のデータを主として送信するため、受信側端末20bにおいて音声再生した際に電話口に入る外界の音などの背景音はほとんど再生されない。このような事態が不自然であると考えられる場合は、あるいは話者の好みに応じて、背景音に関する音源コードを送信し、背景音と話者の音声とを重畳して再生する。あるいは、音楽、電車音等を背景音源コードとして送出してもよい。以上のことは、図2の音声認識・合成変換部2に上記機能を追加することにより実現可能である。

#### 【0068】一第6実施形態一

音声による再生を避けたい会議室などにおいては、音声認識・合成変換部2から直接文字データを出力してデータ蓄積・表示部7でこの文字データを文字出力・表示することが可能である。また、通信制御部3の機能の一部として、データ蓄積・表示部7を用い応答・蓄積機能を備え、留守のため受信側の話者が応答できない場合、通信制御部3が自動応答し、相手側の音声と本人の音声プロファイルデータとを用いてボイスメールデータを作成し、これをデータ蓄積・表示部7に蓄積する事が可能である。あるいは、文字データを電子メール形式で蓄積する事も可能である。さらには、このデータ蓄積及び応答機能を利用し、同様に受信側話者が留守のため応答できない場合、通信制御部3を用いて文字データを電子メールとして、あるいは必要時は音声プロファイルデータを加えたボイスメールとして端末使用者の外出先等に転送する事も可能である。

#### 【0069】一第7実施形態一

カメラ及び画像処理及び表示ソフトウェアをデータ蓄積・表示部7の機能として追加し、話者の顔等の映像データを転送し音声と共に映像を相手側ディスプレイに表示する事も可能である。

#### 【0070】

【発明の効果】本発明は上述のように構成されたので、非ギャランティー型ネットワークを介した通話方法において、通信の不安定性や負荷変動に起因する音声劣化、言葉の不自然な分断、音飛び等を極力排して理解容易な



会話を可能にすると共に、自然な会話を可能にすることができる。

【図面の簡単な説明】

【図1】本発明の第1実施形態である音声送信装置・受信装置を示すブロック図である。

【図2】本発明の第2実施形態である音声送信装置・受信装置を示すブロック図である。

【図3】図2に示す音声認識・合成変換部のブロック図である。

【図4】第2実施形態の動作を説明するためのフローチャートである。

【図5】第2実施形態の動作を示すデータタイムチャート図である。

【図6】第2実施形態及び第4実施形態におけるデータ送信手順を示す図である。

【符号の説明】

1 音声入出力部

1 a マイクロフォン

1 b スピーカ

2 音声認識・合成変換部

2 a 音声認識回路

2 b 圧縮符号化回路

2 c 選択回路

3 通信制御部

4 網制御部・アクセス部、網制御装置

5 モデム

5 a パケット送信および通信制御部

5 b パケット受信および通信制御部

6 入力操作部

7 データ蓄積・表示部

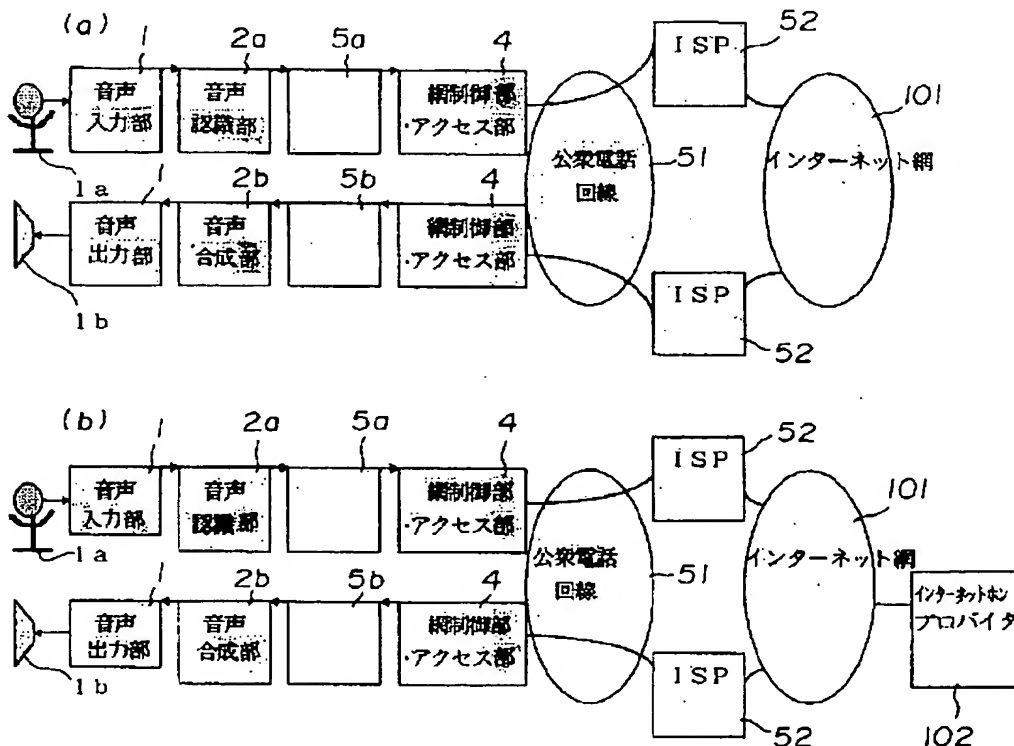
5 1 公衆電話回線

5 2 インターネットサービスプロバイダ

1 0 1 インターネット網

1 0 2 インターネットホソサービスプロバイダ

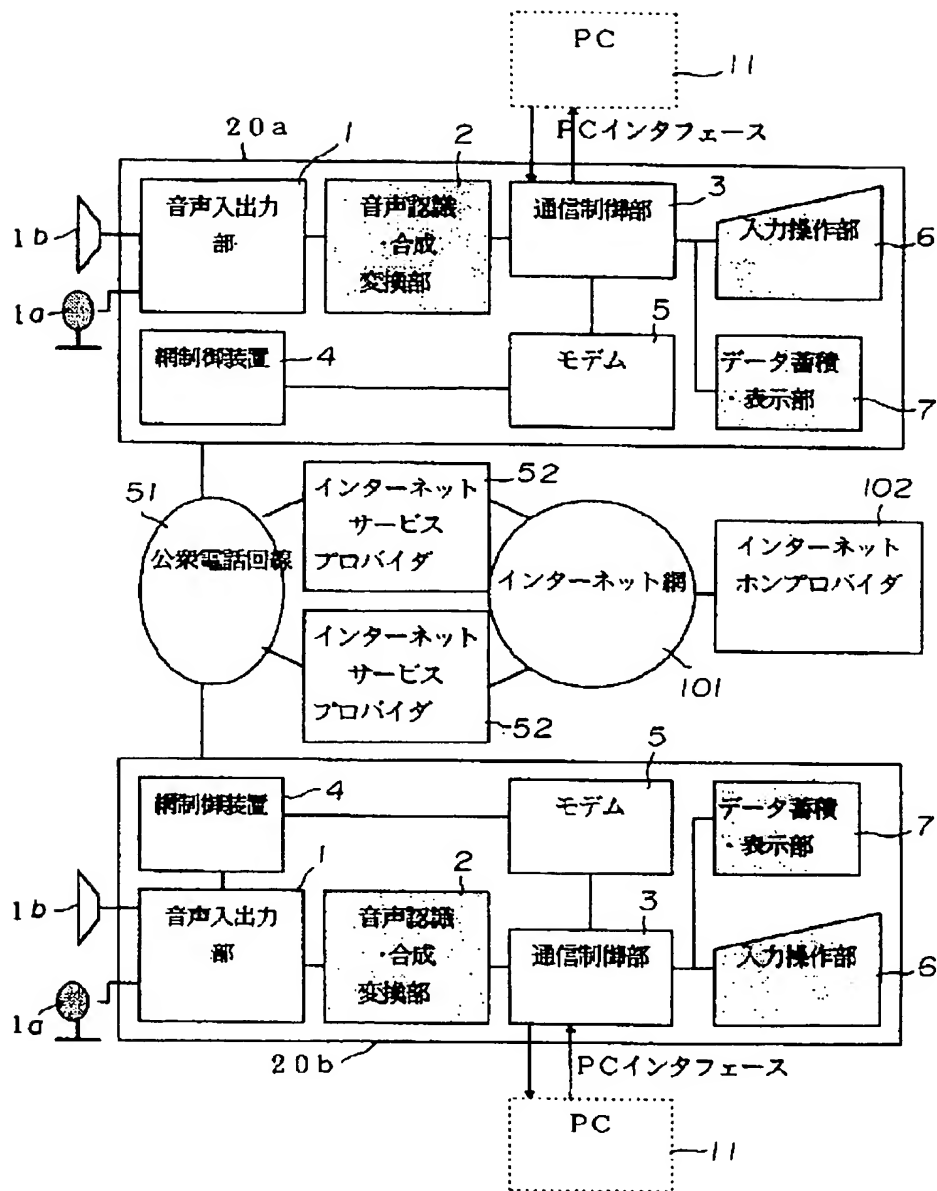
【図1】



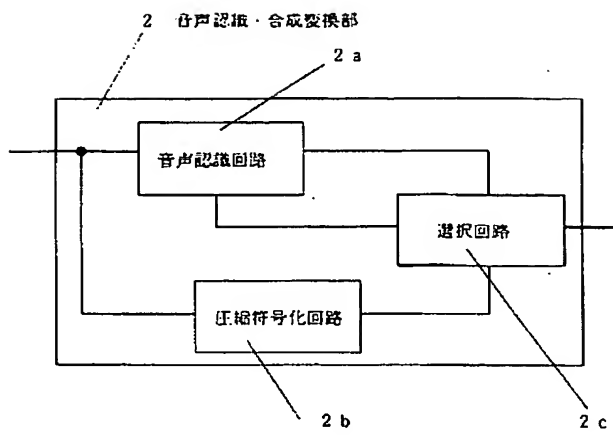
5 a : パケット送信および通信制御部

5 b : パケット受信および通信制御部

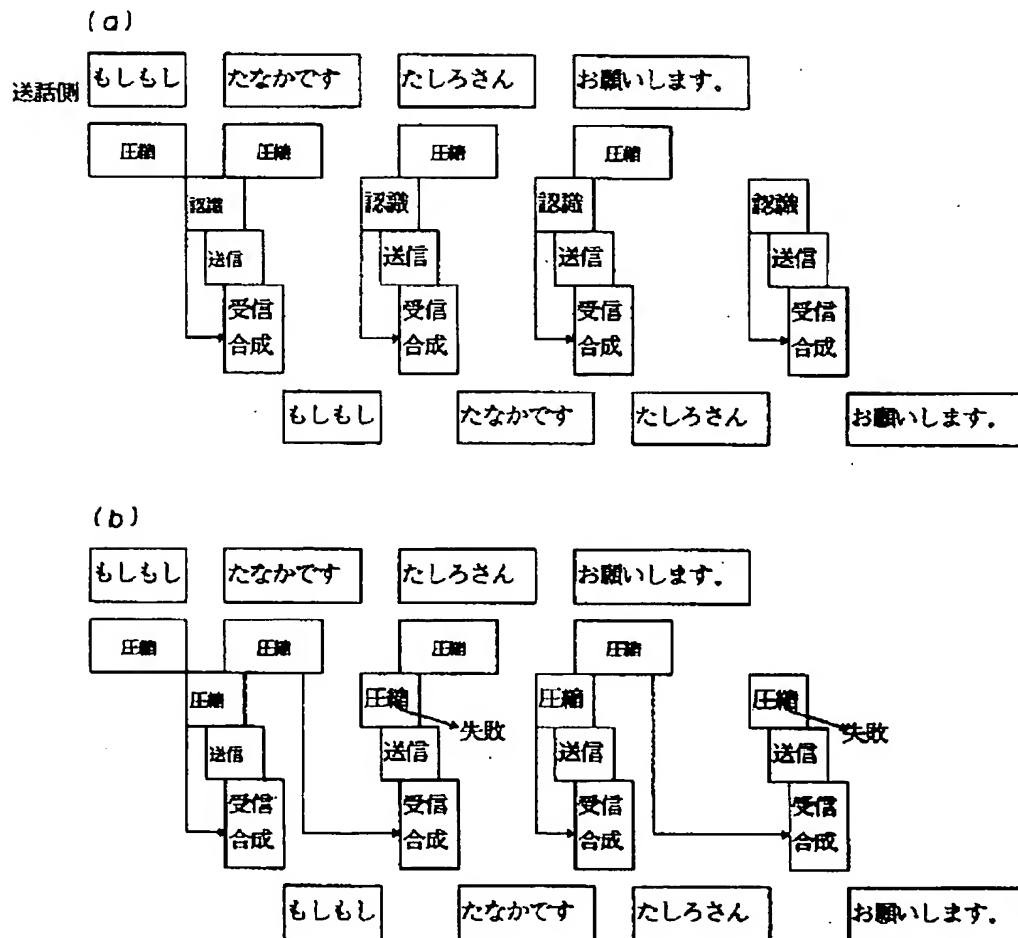
【図2】



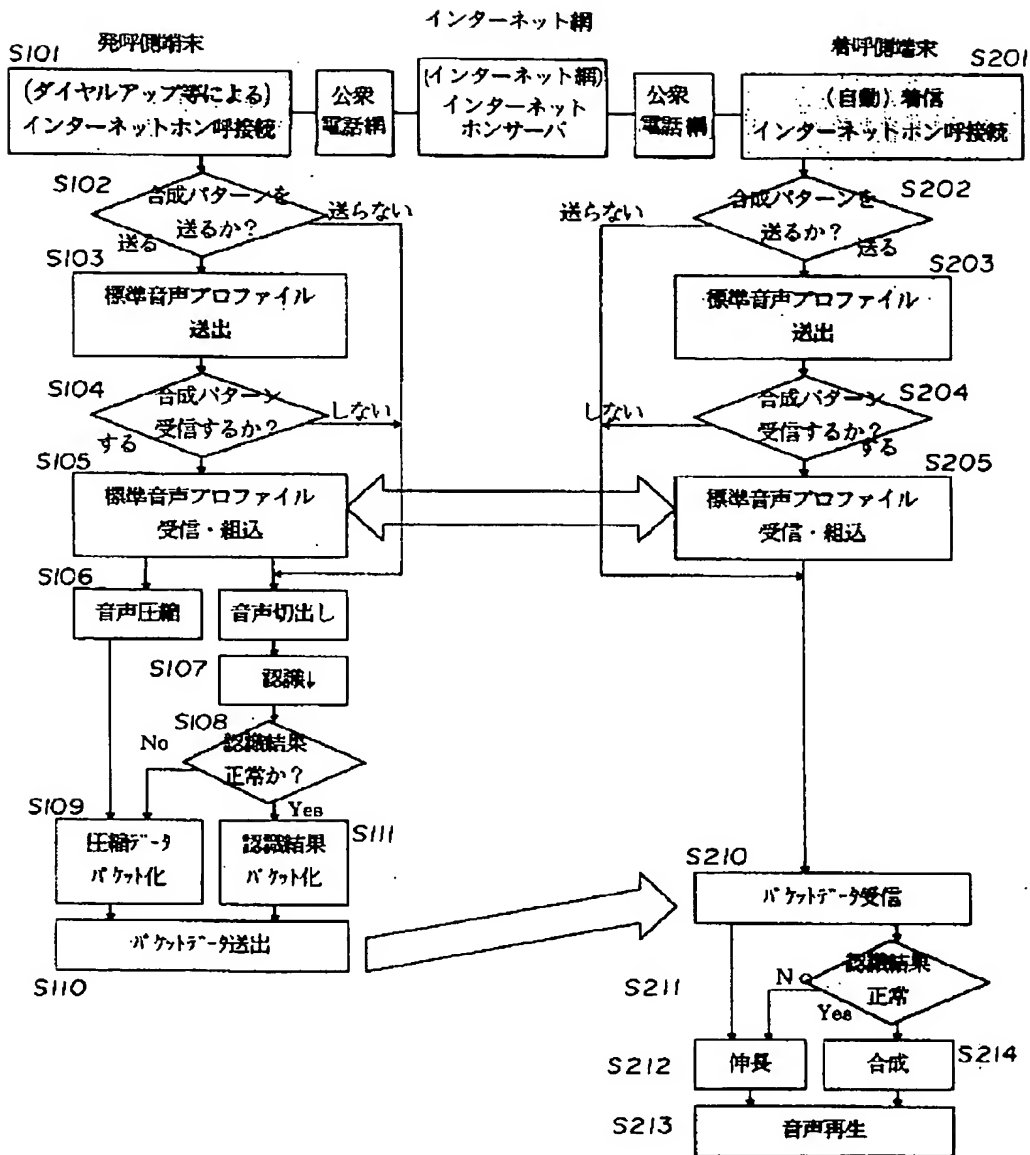
【図3】



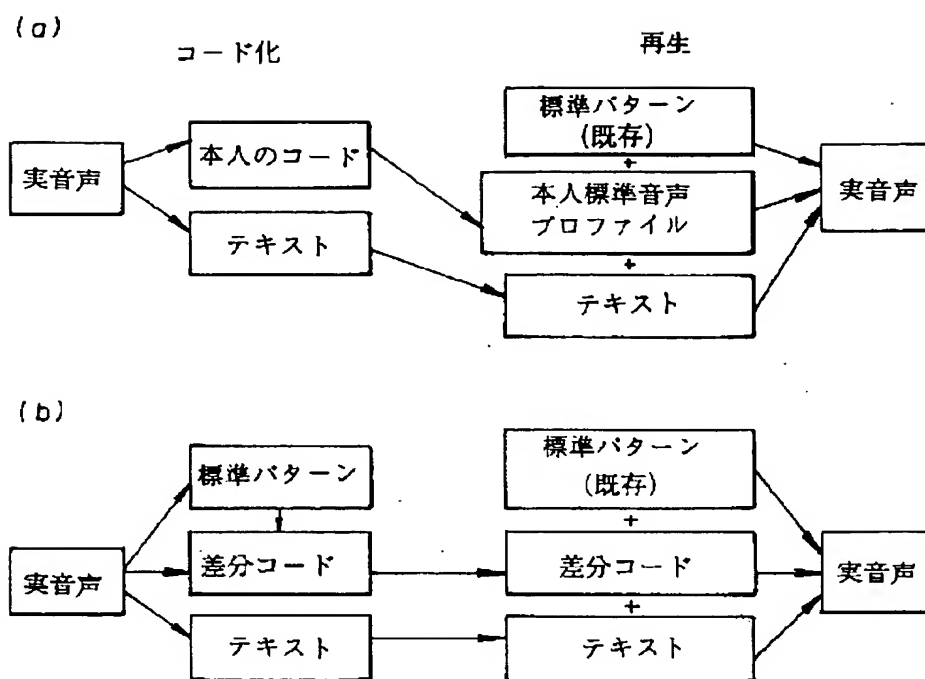
【図5】



【図4】



【図6】



フロントページの続き

(51) Int. Cl.<sup>6</sup>

H04Q 11/04

識別記号

F I

H04Q 11/04

R